

34	168	3.1	3924	1	ANK2_HUMAN
35	167	3.1	1639	1	MSPI_PLAFW
36	165.5	3.1	1256	1	ATL_STAAU
37	164.5	3.0	1164	1	BAG_STRAG
38	164.5	3.0	1189	1	YJH6_YEAST
39	164	3.0	1222	1	S160_YEAST
40	164	3.0	1679	1	YMF9_YEAST
41	163.5	3.0	1528	1	SPAA_STRDO
42	163.5	3.0	1545	1	IGA3_HAEIN
43	163	3.0	1159	1	N124_SCHPO
44	163	3.0	1460	1	N159_YEAST
45	162.5	3.0	857	1	NFM_CHICK

Q01484 homo sapien  
P04933 plasmodium  
P52081 staphylococ  
P27951 streptococc  
P47035 saccharomyc  
P06105 saccharomyc  
Q04958 saccharomyc  
P21979 streptococc  
P45385 haemophilus  
Q09904 schizosacch  
P40477 saccharomyc  
P16053 gallus gall

## ALIGNMENTS

RESULT 1  
MSPI\_PLAFM  
ID MSPI\_PLAFM STANDARD; PRT; 1701 AA.  
AC P08569;  
DT 01-AUG-1988 (Rel. 08, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE MEROZOITE SURFACE PROTEIN 1 PRECURSOR (MEROZOITE SURFACE ANTIGENS)  
DE (PMMSA) (P190).  
GN MSP-1.  
OS Plasmodium falciparum (isolate mad20 / Papua New Guinea).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=70153;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88011243; PubMed=3079521;  
RA Tanabe K., Mackay M., Goman M., Scaife J.G.;  
RT "Allelic dimorphism in a surface antigen gene of the malaria parasite  
Plasmodium falciparum.";  
RL J. Mol. Biol. 195:273-287(1987).  
RN [2]  
RP REVISIONS TO 1403; 1569 AND 1629.  
RA Tanabe K.;  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 1-115 FROM N.A.  
RX MEDLINE=86136024; PubMed=3004972;  
RA Mackay M., Goman M., Bone N., Hyde J.E., Scaife J., Certa U.,  
RA Stunnenberg H., Bujard H.;  
RT "Polymorphism of the precursor for the major surface antigens of  
Plasmodium falciparum merozoites: studies at the genetic level.";  
RL EMBO J. 4:3823-3829(1985).  
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR  
(POTENTIAL).  
CC -1- PTM: MEROZOITE SURFACE ANTIGEN CONTAIN THE SEQUENCE OF 83 KDA, 42  
KDA AND 19 KDA ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF  
MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X05624; CAA29112.1; -;  
DR PIR; A26868; A26868.  
DR PIR; B25120; B25120.  
DR InterPro; IPR000561; -;  
DR Pfam; PF00008; EGF; 1.  
KW Malaria; Merozoite; Polyprotein; Repeat; Signal; Glycoprotein;  
KW Transmembrane; GPI-anchor.  
FT SIGNAL 1 19 POTENTIAL.  
FT CHAIN 20 1701 MEROZOITE SURFACE PROTEIN 1.  
FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 239 239 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 470 470 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 536 536 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 607 607 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 802 802 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 899 899 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 919 919 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 965 965 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 991 991 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 1089 1089 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 1196 1196 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 FT CARBOHYD 1588 1588 N-LINKED (GLCNAC. . .) (POTENTIAL)  
 SO SEQUENCE 1701 AA: 193768 MM: 3FC2EC59AF96AE98 CRC64;

Query Match 4.28; Score 227.5; DB 1: Length 1701;  
 Best Local Similarity 20.2%; Pred. No. 0.005; Indels 405; Gaps 66;

Matches 250; Conservative 176; Mismatches 405; Indels 405; Gaps 66;

1 MKFSKYYIAGSAVYSLCAVALNHRSEEN---KDNRRVYVDSQSSQSKSENLP 56  
 500 MKFNNNF---DKDVYDKIFSAITYNEKORYNNKSSNSNYYV---QKLKALSTLE 553  
 57 DVSQKCEIOAEQIVIKITDQGYVSHGHYHYNGKVPYDA---LPSEEL----- 104  
 554 D-YSLRKGIT-----SEKDFNHYITLKTGLEADIKKLEIKSEKSEKILE 596  
 105 -----LMQDPYQKLDADYV-EYKGGYII-KYDG--KYVYVYLKDAHADNVNT----- 149  
 597 KMFGLTHSANALESVSDIYKLOVQKYLKIKIEDLKIEFLKNOLOKDSIHVPNIYKP 656  
 150 -----KDEINROKQ-----EHVKNDEKNVSNVA---VARSQGRYTTNDG--- 185  
 657 QNKPEPYLYLKLKEVEVKKEFIPKVKMDLKKQAVLSSITQPLVAASE---TTEDGSHS 713  
 186 -----YFNPADIIEDTCNATYVPHGCHYHYI-----PKSLSASELAAKAHILA 230  
 714 THTLSQSGEFTVTEETVTEET-----VGHITVTITLPRKESAPKEKVENSIIE 765  
 231 GKNMPSQLSSYSTASDNNOOSVAKGSTSKPANK-----SENLOSILKEL 275  
 766 HK-----SNDNSQALTKTYLKKLDEFLTASYICHKYLIVSNSSMDOKILEV 812  
 276 YD-SPSAQRYSESQDLVDFPAKIIISRTPNGVALPHG-----DHYHPIPYSK----- 320  
 813 YULTPEEKEKLS-----CDPLDLLFNIONJMPAYSLYDSMNNDLQHLFPFLYQKEMITY 868  
 321 LSALEEK-----IARNVPLISGTSTVST--NAKPEVYVSSGLSSNPSLTTSKELSSA 373  
 869 LHLKEENHIKLLLEEQKOLITGSTSTSPGNTVYNTAQSATHSNOSNOOSNASS---TNT 925  
 374 SDGYIFNKKDIVEETATYIVRHGDHFHYIPKSNQI--GQPTLPN--NSLATPSPSLP 430  
 926 QNG-----VAVSSGPAVVEESHDPPLTVLISNDLGIYSLNLGKTKYVPPN-LLTST 977  
 431 GTSHEKHEDYGFEDANRIIADESGFVMSGHGHNHYFFRKD-----LTEQIKAAQKH 484  
 978 -TEMEKFEYEN-----ILKNNDYF-----NDIKOFVKSNSKVITGLTETQKNALNDE 1024  
 485 LEEVYTS-----HN-----GLDLSSEHODYPQNAKEMD---LKKITEKTAGIMKOY 530  
 1025 IKKLDYLOLSFDLYNKKYLKLDRLFNKKKELGODKMQIKKLLLEQLESKLSLNNPH 1084  
 531 GYKRE-STIVNKKENKATII-----YPHGDHHAADPIDENKHPVGI 567  
 1085 NVLONPVYFPNKKKEAEAELENTLENTKILKHYKGLVYKYNCE---SSPLKTLSEVSI 1141  
 568 --GSHSHYVELFKPEGVAKKEGKNKYTGEE-----LTNVVNLKKNSTENNQNTLA 617  
 1142 QTEQDNYANLEKFRALSKIDGLTNDNLHGLKKKLSFLSSGLHLITELK-EVIKKNKNT-- 1198  
 618 NGQKNVSSFPPELEKKGITMLYKL---ITPDQKYLEKVSQKVFGEVGNINIANFELDOP 674

DB 1199 -----GNSPSENNKK--VNEALSKYENFFPEAKVTTVTP-----POP 1234  
 QY 675 YLPQGTFFRYTI-----ASKDYPEVSYDGTFTVPTSLAKMASQTIFFPHAGDLYLRNP 729  
 DB 1235 DVTPEPLSVRVSGSSGSKKEETOIPTSG-----SLTELOOVVOLQNYDEEDDSLVLPP 1288  
 QY 730 QPAVFKGD-----ALVRVDFEFGNAYLENNKYGEIKLPKLNQGTTRAGN 779  
 DB 1289 IFGESEDEYLDQVVTGEALSVTMDNLSG--FENEVDIYIK-PL---AGYYRSLKK 1341  
 QY 780 KIP-----VTFMANA-----YLD-----NOSTYIVVPILEKENOTD 811  
 DB 1342 QLEKNIITFNLNILNLSRLKKRKYPLDYLESMDLOFNHISSENYIE--DSFKLLNSEQ 1400  
 QY 812 KPSILPQFRNK-----AOENSKLDEKY-----FEPTSEKVEKEXLSEGENS 854  
 DB 1401 KNTLLSKYKYEKESYENDIKFQOEGISYEEKVLAKYKDDLESIKKVIKEKEKPPSSPT 1460  
 QY 855 TNSNLEEVPTVDPQOEVAKRA-----ESYCKKLEVLNFM----- 891  
 DB 1461 TPEPS-----PAKTDQKESKFLPPLTNITETLYNHLVKNKIDYILNLAKAKINDCNEXKD 1515  
 QY 892 -----DGTIELYLPSEGVIKKNMADEFTGEAPQOGENKPESENCK-VSTGTV 936  
 DB 1516 AHVKITKLSDLKAIDKIDLF-----KNTNDFEAIKKILINDDTKKDKMLGLSTGLV 1567  
 QY 937 ENQPT-----ENKPADSLPEAPNE--KPVAPENS 963  
 DB 1568 QNFPPTIISKLEGGFQDMLNISOHCYVKKOCPENS 1603

QY	6	LCAAYALNHOHSGONKDNRRVSYDGSOSKSS	-----	ENTLPPQVS	-----	QKEIOAEQIVT	58
Db	170	LCAYAAANONLSQLGMR	-----	VASNKLCHTFRANTPAPADIDLSRGTAYEYVLH	22		
QY	59	KIRIDOGYVSHSDHHYHYNG	---KVPIDALFSEELMKOPNTQLDADIVNEVKGYYIK	115			
Db	225	DLENQVIL	-GGRHLEELGEVKKPALDADLISMTGMTKCDALENVSILTEETIOEGTVK	283			
QY	116	YDGKTYVYLKDAAHADN	132				

ENTRY	DATE	RESULT
1943.METUA	9	
1943.METUA		
1943.METUA		
15-JUL-1998 (Rel. 36, Created)		
15-JUL-1998 (Rel. 36, Last sequence update)		
15-JUL-1998 (Rel. 36, Last annotation update)		
HYPOTHETICAL PROTEIN MJ0943.		

Db 235 ALSQRYKONFENILENKIVNIPVGGKKNVAVRVAEFDSQKSKTETFDPPNNVYR 294  
 QY 46 -----SQKEGIAEQIVKIDTQGYVTSHGHHYHYNGKVPYDALFSEELMKDPNYOL 99  
 Db 295 SEONLEKEGAPKKQVG---GVPGVSVNIGPVQGLKDNKEPE---KYEKSONTNTIEV 347  
 QY 100 KDADIVNEVKGY-----IIRKVDGKYVYVYLKDAANA 130  
 Db 348 --CKTISEIKGFGTLVRLNAVAVVGVGKYKTALEDGANA 384  
 RESULT 6  
 14  
 C:Species: Plasmodium falciparum  
 C:Date: 13-Nov-1998 #sequence\_revision 13-Nov-1998 #text\_change 21-Jul-2000  
 C:Accession: D71614  
 R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;  
 :; Perle, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.  
 Science 282, 1126-1132, 1998  
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.  
 A:Reference number: A71600; MUID:99021743  
 A:Accession: D71614  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-2573 <GAP>  
 A:Cross-references: GB:AE001396; GB:AE001362; NID:g3845188; PIDN:AACT1881.1; PID:g384519  
 A:Experimental source: clone 3D7  
 C:Genetics:  
 A:Gene: PFB0460C

Query Match 11.2%; Score 79.5; DB 2; Length 2573;  
 Best Local Similarity 20.8%; Pred. No. 1.2e+02;  
 Matches 31; Conservative 25; Mismatches 38; Indels 55; Gaps 6;  
 QY 12 NQHRSEKNNRNVYDGSQSQSE-----NLTPD-----OV 45  
 Db 2421 NDNNDNNNDNNNSHTAFQNTQGETTNNITNDICEKGNKYSNNVNNINEM 2480  
 46 SQKEGIAEQIVKIDTQGYVTSHG---DHVY-----YNGKVPYDA 84  
 LQ 2481 TKCESVEVNEIIQTNKRKF---HNIELKEHYCYDLFKRKLNTYRTYKKNRIINC 2537  
 QY 85 LFSSEELMKDPNYOLKADIVNEVKGYI 113  
 Db 2538 LITNKNI-----FOYKEHDIWVKVQIFI 2561

RESULT 7  
 JCA109  
 triacylglycerol lipase (EC 3.1.1.3) 1 - Mycoplasma mycoides subsp. mycoides  
 N:Alternate names: lipase  
 C:Species: Mycoplasma mycoides subsp. mycoides  
 C:Date: 23-Jul-1995 #sequence\_revision 19-Oct-1995 #text\_change 07-Dec-1999  
 C:Accession: JCA109  
 R:Rawadi, G.; Lalanne, J.L.; Roulland-Dussoix, D.  
 Gene 158, 107-111, 1995  
 A:Title: Cloning and characterization of the lipase operon from Mycoplasma mycoides subsp.  
 A:Reference number: JCA109; MUID:95509706  
 A:Accession: JCA109  
 A:Molecule type: DNA  
 A:Residues: 1-262 <RAW>  
 A:Cross-references: GB:U17036  
 A:Note: The authors translated the codon TGA for residue 162 and 236 as Trp  
 C:Comment: This enzyme, a serine esterase, is widely distributed throughout animals, pla  
 tly acids, and hydrolyses ester bonds of triacylglycerols to yield free fatty acids, dig  
 C:Genetics:  
 A:Genetic code: SGC3  
 C:Superfamily: triacylglycerol lipase 1  
 C:Keywords: carboxylic ester hydrolase  
 F:90-94/Region: conserved site

Query Match 11.0%; Score 78; DB 2; Length 262;  
 Best Local Similarity 26.4%; Pred. No. 9.9;  
 Matches 29; Conservative 15; Mismatches 50; Indels 16; Gaps 4;  
 QY 9 YALNHRSEKNNRNVYDGSQSQSENLTPDQVSKQEGIAEQIVI--KITDQGYV 66  
 Db 5 YDYNVYFKNNNDNNENIIFVHYNSS-----PTPEYLNKIQDQILIMHYNFDDQIYV 57  
 QY 67 TSHGDHYYHYNGKVPYDALFSEELMKDPNYOLKADIVNEVKGYIIV 116  
 Db 58 KPVKDH-----KVTVEG-FAQLLIHFIEQNOIKNVVAIGHSMGSGVISI 100  
 RESULT 8  
 I40824  
 spore cortex-lytic enzyme precursor - Clostridium perfringens  
 C:Species: Clostridium perfringens  
 C:Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 15-Oct-1999  
 C:Accession: I40824; PC2363  
 R:Miyata, S.; Moriyama, R.; Miyahara, N.; Makino, S.  
 Microbiology 141, 2643-2650, 1995  
 A:Title: A gene (slec) encoding a spore-cortex-lytic enzyme from Clostridium perfring  
 A:Reference number: I40823; MUID:96036223  
 A:Accession: I40824  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-438 <RES>  
 A:Cross-references: GB:DA5024; NID:g940391; PIDN:BAA08081.1; PID:g940393  
 R:Miyata, S.; Moriyama, R.; Sugimoto, K.; Makino, S.  
 Biosci. Biotechnol. Biochem. 59, 514-515, 1995  
 A:Title: Purification and partial characterization of a spore cortex-lytic enzyme of  
 A:Reference number: PC2363; MUID:95252603  
 A:Accession: PC2363  
 A:Molecule type: protein  
 A:Residues: 150-164 <MIY>  
 C:Genetics:  
 A:Gene: slec

Query Match 11.0%; Score 78; DB 2; Length 438;  
 Best Local Similarity 28.6%; Pred. No. 18;  
 Matches 26; Conservative 10; Mismatches 27; Indels 28; Gaps 4;  
 QY 58 KITDQGYVTSHGHHYHYNGKVPYDAL---FSEELMKD-----PNYQLKDA- 102  
 Db 278 INVQCPGWTQWGSKYLGDEGKVPYDILTSTFYGDLELAKSAKKVGSPPSYPGYTLTKGY 337  
 QY 103 -----DIVNEVKGY----IIRKVDGKY 120  
 Db 338 SGPEVRVQIQLNAISRAYPLIPKIAVDGKY 368  
 RESULT 9  
 DB1401  
 probable flagellar hook-associated protein Cj0548 [imported] - Campylobacter jejuni  
 C:Species: Campylobacter jejuni  
 C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 31-Mar-2000  
 C:Accession: DB1401  
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chil  
 Nature 403, 665-668, 2000  
 C.W.; Quail, M.; Rajadream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Ba  
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals  
 A:Reference number: A81250; MUID:20150912  
 A:Accession: DB1401  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-642 <PAR>  
 A:Cross-references: GB:AL111168; NID:g9697817; PIDN:CAB75184.1; PID:g9596  
 C:Experimental source: serotype O2, strain NCTC 11168  
 C:Genetics:  
 A:Gene: flid; Cj0548